Data Science Intern at Data Glacier

1. Project: Bank Marketing

Week 10: Deliverables

Name: İlke Candan Bengi

University: Marmara University

Email: ilkecandan@hotmail.com

Country: Turkey

Specialization: Data Science

Batch Code: LISUM12

Date: 06 October 2022

Submitted to: Data Glacier

1-Problem Statement

The bank wishes to offer its term deposit product to clients, and before doing so, they want to create a model that will assist them determine if a certain consumer would buy their product or not (depending on the customer's previous interactions with the bank or another financial institution). Bank wants to utilize ML model to shortlist consumers who are more likely to buy the product so that their marketing channels (telemarketing, SMS/email marketing, etc.) may focus exclusively on those customers who are more likely to buy the product.

2. Cleansing and modification of data

First and foremost, we remove all relevant data. Any rows with missing values will be removed. As you can see, the number of rows has shrunk. The entire trait is then removed. This deletes the whole feature/attribute. Consider the following instances. The age column is no longer present. We give the missing values a value. If you feel the attribute is important enough to include in training. You are free to fill in the gaps. You can use the median, mean, or zeros to fill in the missing data. Any missing values will be replaced by the calculated median.

You'll also see that the mean value has changed once we've filled the property.

3- EDA Performed on Data

Firstly, we took a look at how the “balance” factor is distributed. It is the most important element of our data. Then, we perform numerical data distribution. Then, for this part lets look at the distribution of all of the features by ploting them.To do so, we first list all the types of our data from our dataset and take only the numerical ones.

Then, we check which factors are mostly correlated with the balance factor. The factors of duration, pdays, day and campaign were the mostly correlated ones.